ChemPass

SynSpace

Multistep forward synthesis for scaffold hopping and generative design in synthetically feasible chemical space

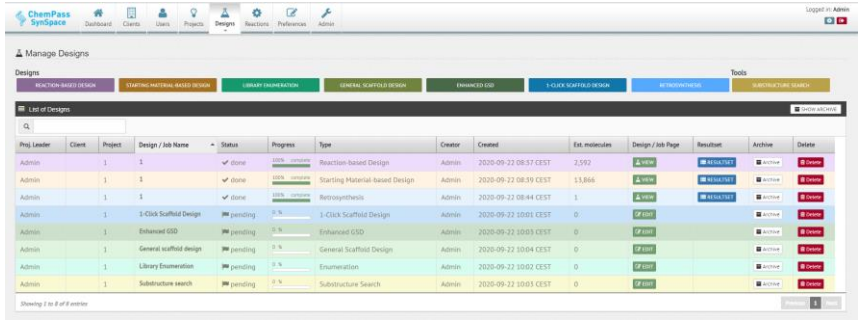
NIH Virtual Workshop on Reaction Informatics 2021

ChemPass

- ChemPass:
  - A platform company service provider in small molecule preclinical space
    - Medicinal chemistry-driven technology development
  - Founded in 2016
- ChemPass platforms
  - Highly validated: >45 lead optimization collaborations to date

ChemPass SynSpace

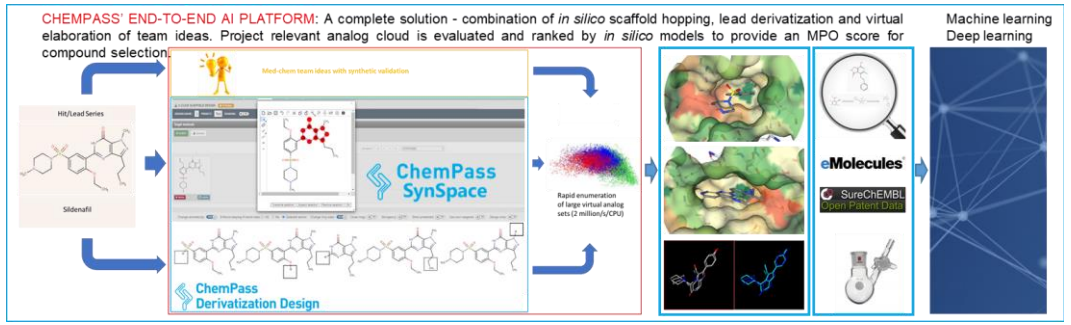
ChemPass Derivatization Design



Design software for medicinal or computational chemists

ChemPass SynSpace

ChemPass Derivatization Design



End-to-end AI-assisted lead Discovery (AID) platform

Key considerations impacting preclinical cost and timelines

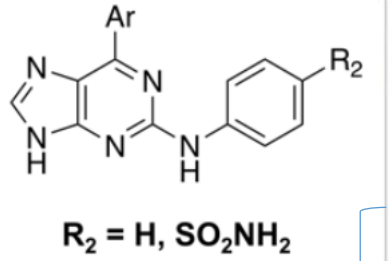
- Historical lead discovery process at Big Pharma
  - 2.5-4 years, 250-450 FTE months spent
- DMTA cycle time and the # cycles are critical factors

- Synthesis is far the longest and most expensive
  - 0.5 – 4 months per cycle
  - Data of stragglers miss out on several cycles
  - Mindset problem:
    - Compelling design ideas are hard to dismiss due to perceived synthetic challenges
    - Hard to drop a compound once synthetic effort is underway as there is always “another route”
- Off-the-shelf and virtual commercial catalogs will have <u>little impact after lead finding</u>

eMolecules® REAL DATABASE SAVI

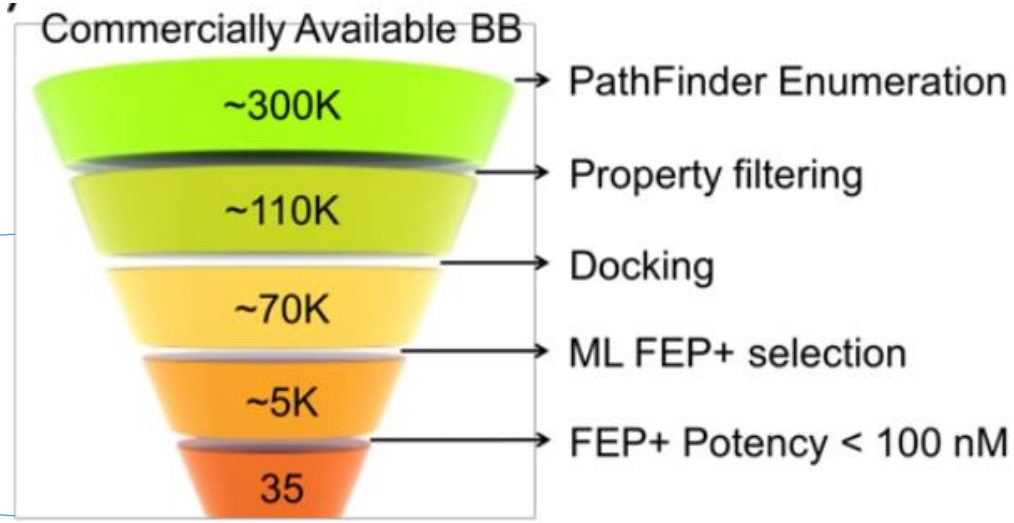
Where to take a hit?

- A suggested computer-aided lead optimization workflow (doi.org/10.26434/chemrxiv.14153819.v1):
  - Either spending a lot of money in an expensive process to carry compounds that get thrown out by retrosynthesis analysis
  - Or spending a lot of money to run retrosynthesis on 100K-millions of compounds, most of which get thrown out in the docking/FEP stages: perhaps RAscore can help (DOI: 10.1039/d0sc05401a)



R_2 = H, SO_2NH_2

Resource intensive:
- Cloud cost
- License cost



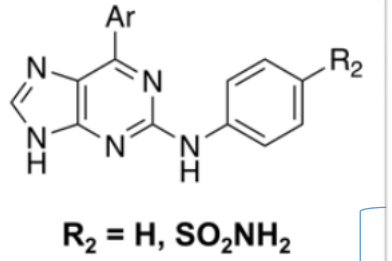
Full retrosynthesis here?
- Cloud cost
- License cost

Full retrosynthesis here?
- Cheap

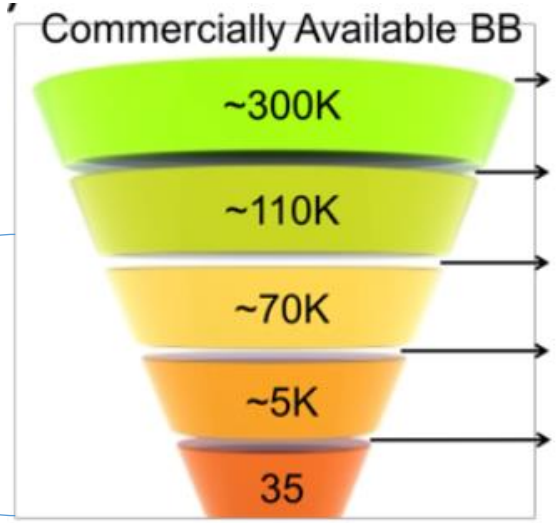
In real projects, typically we have a much bigger chemical space: **30-70% of the resources and cost will be wasted**

A much better process

- What if the **entire idea space is synthetically feasible**
  - Costly processes are carried out on relevant synthetically feasible chemical space
  - But is the space as good as offered by deep generative design or simple enumeration?



R_2 = H, SO_2NH_2



ChemPass SynSpace

ChemPass Derivatization Design

Property filtering

Docking

ML FEP+ selection

FEP+ Potency < 100 nM

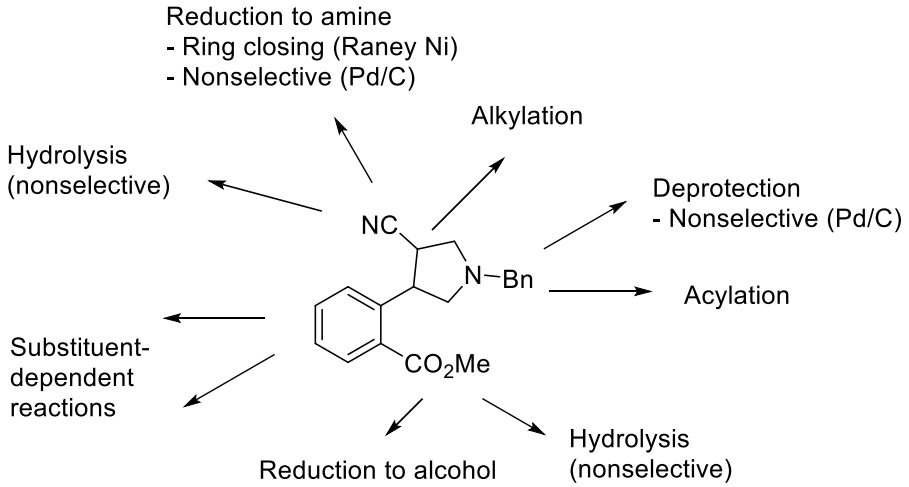
Resource intensive:
- Cloud cost
- License cost

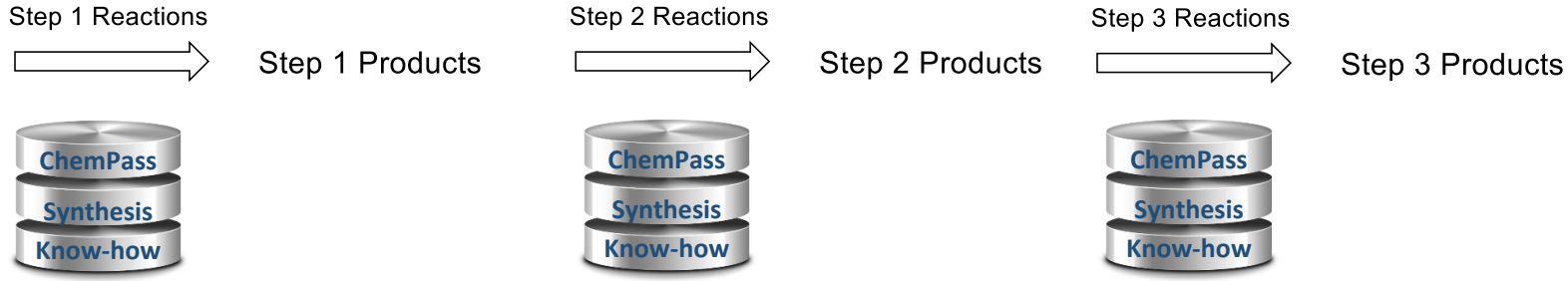
Full retrosynthesis here?
- Cheap

All designed structures are worthy of evaluation: no wasted resources and cost

In silico synthesis as a design concept

- ChemPass Design Technology: What can be synthesized from starting materials, intermediates or lead structures?





Reactivity Assessment

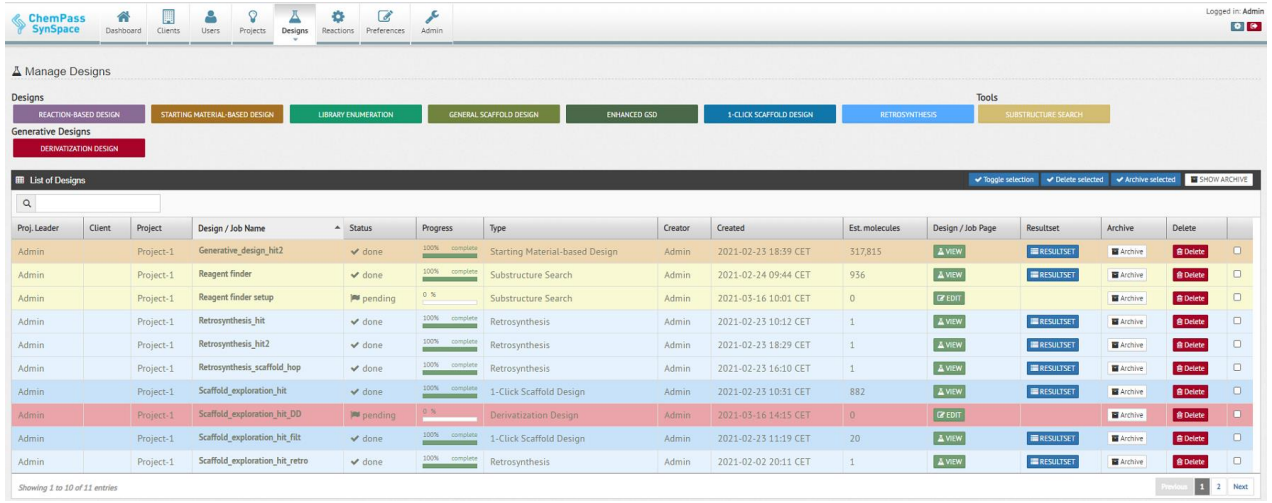
Reactivity Assessment

Reactivity Assessment

- Technology development was required to solve
  - Rule-based AI for forward *in silico* synthesis
  - Molecule design based on multistep *in silico* synthesis
  - Control of combinatorial explosion

SynSpace

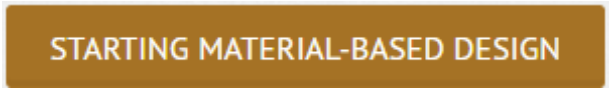
- Custom ideation in synthetically feasible space
  - >300 transformations in current version
- A user-friendly computational tool that can alleviate the boundary between medicinal and computational chemists making preclinical research more efficient
  - No need for synthesis knowledge for computation chemists
  - No need for cheinformatics knowledge for medicinal chemists



SynSpace software: synthesizable chemical space explorations

- Design tasks - SynSpace modules
  - Library design (e.g. DEL)

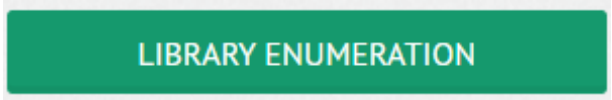
  - Side-chain analog design



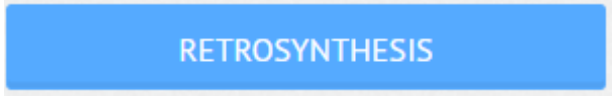
  - Scaffold hopping, scaffold analog design
    - 1-step process: 1-Click design

    - 2-step General scaffold design

  - Multi-step and multi-site library enumeration



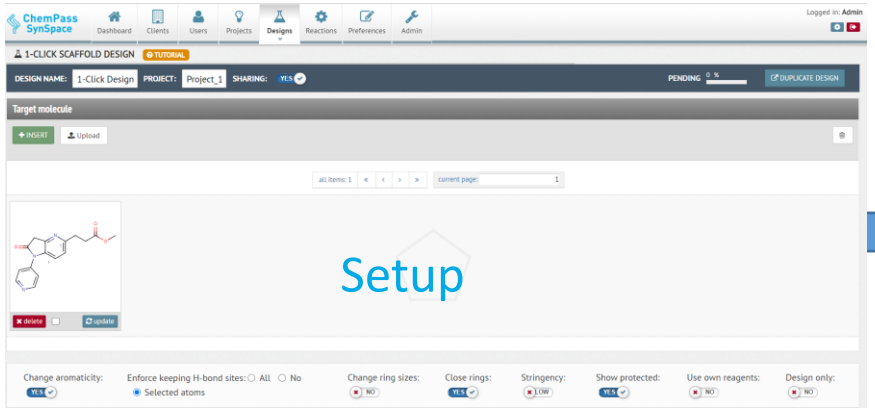
  - Retrosynthesis module

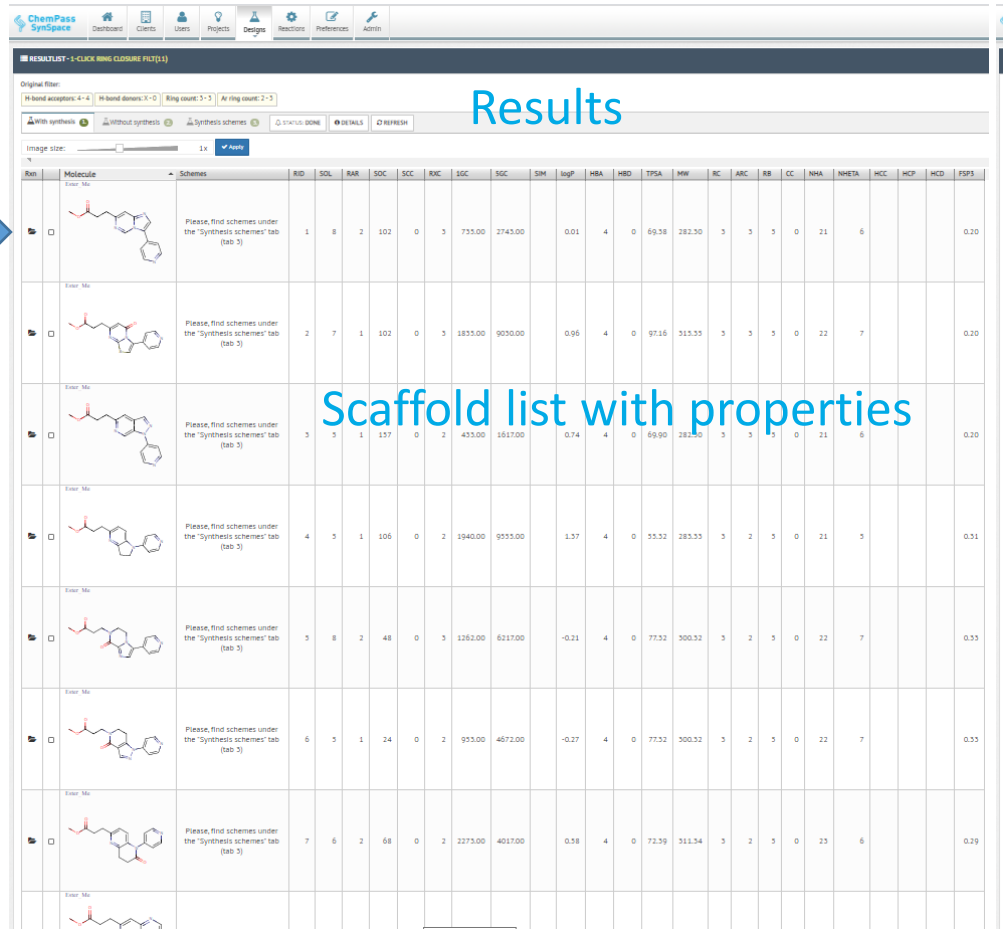


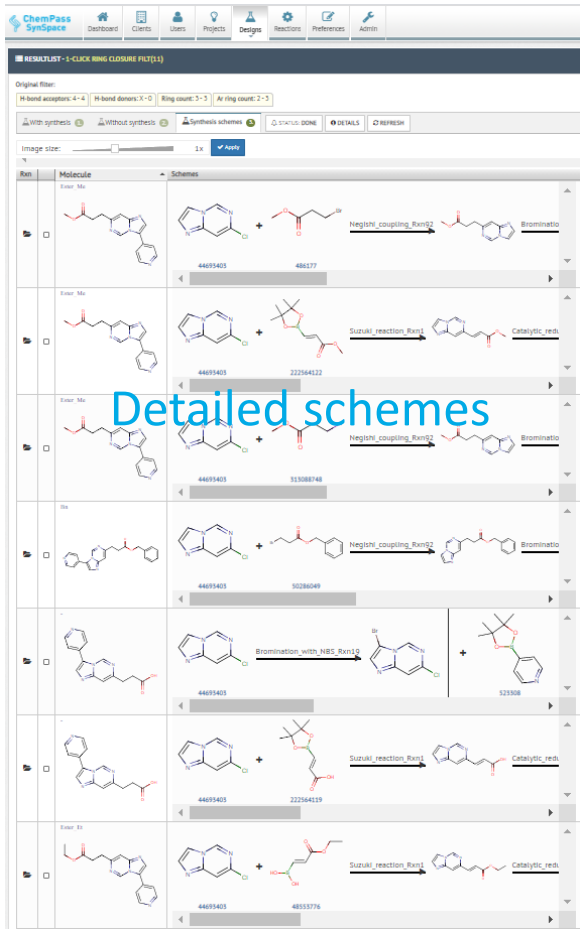
  - Automated generative module: Derivatization design

  - Automated SMBD Generative Design

SynSpace 1-Click scaffold design (forward synth/3D overlap)







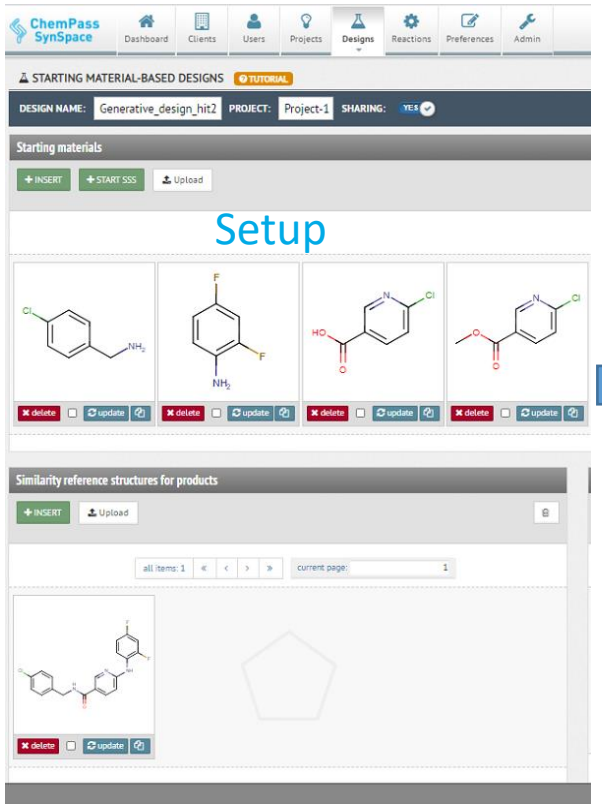
- Simple design tool that requires <u>no cheminformatic skill set</u>
- Design outcome influenced by simple user settings:
  - H-bonding features
  - Aromaticity
  - Ring size
- Intelligent ring closing method included
  - Bicyclic derivatives of monocyclic leads can be easily explored

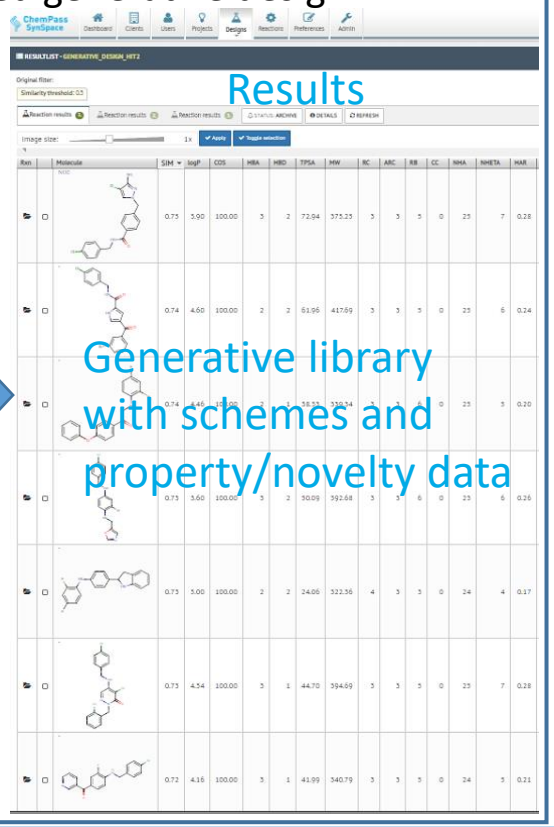
- New scaffolds with properties, synthetic information, novelty assessment

Two proprietary generative design tools

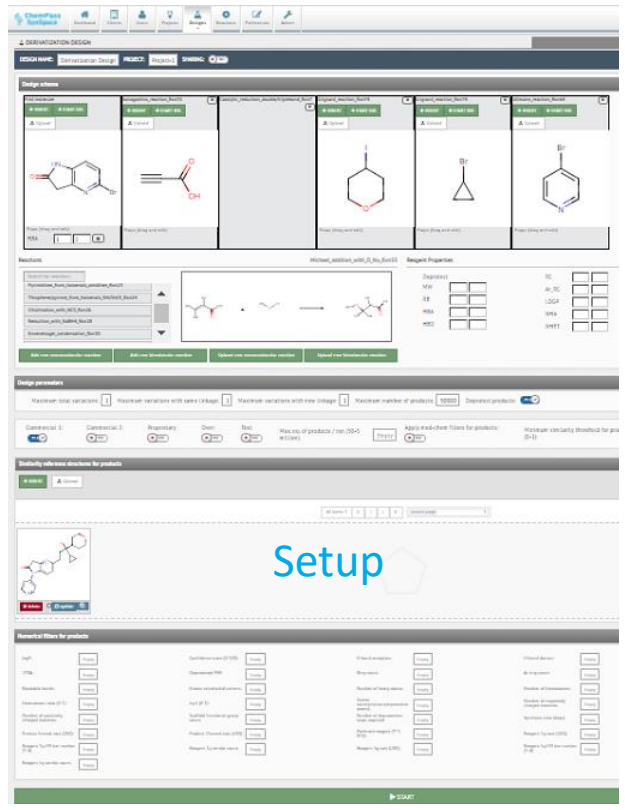
- 2 different user-friendly solutions

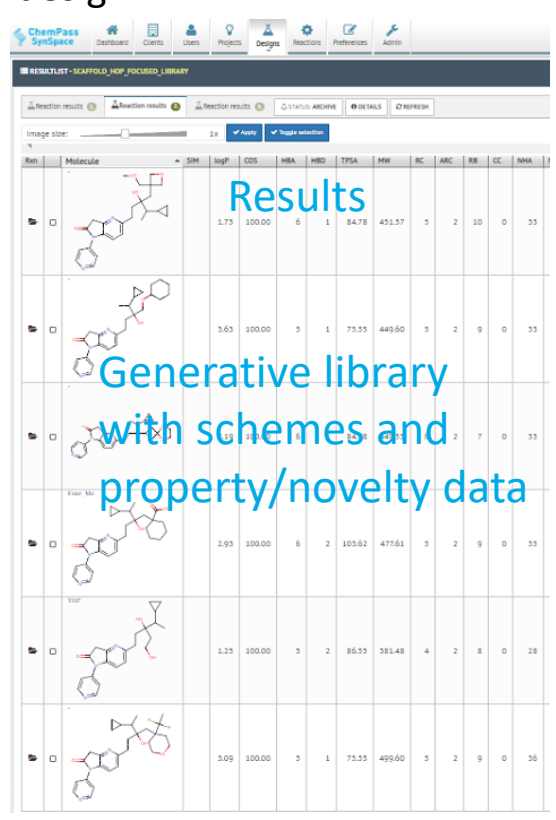
Component (reagent) based generative design





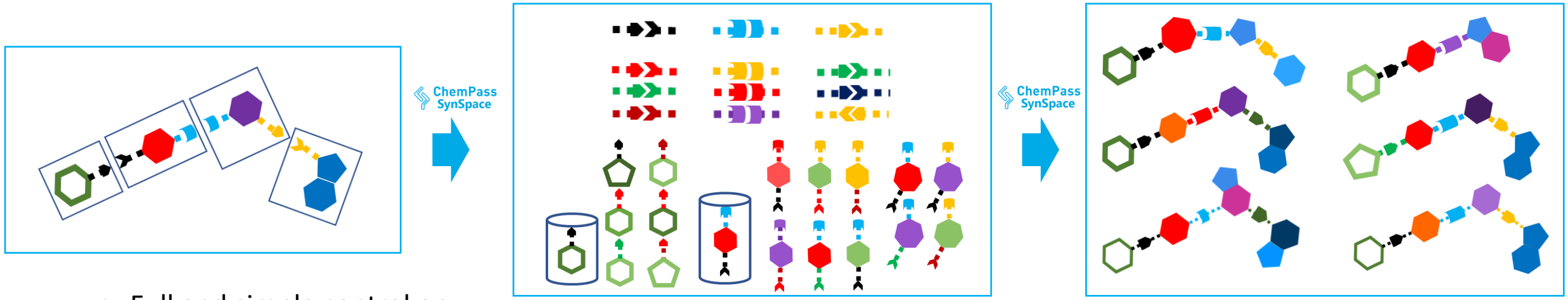
Derivatization design





What is Derivatization Design?

- A universal generative design technique



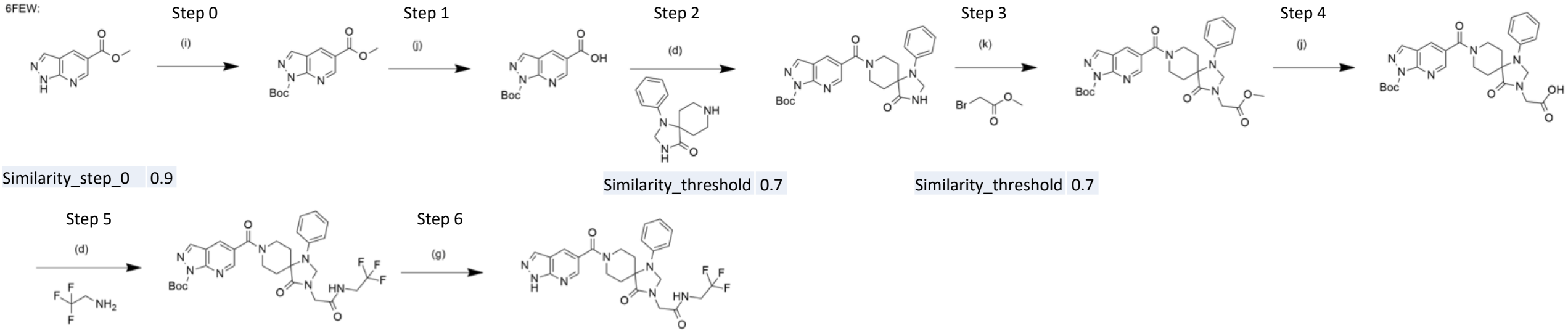
- Full and simple control on
  - # variations, type of variations
  - Depth of modification at each site: scaffold hopping and/or variational analogs (exploration or exploitation)
  - Similarity and desired set size
- Simple user inputs drive the fully automated process

All designed molecules possess vital synthesis, reagent and vendor data

ACS Med. Chem. Lett. **2021**, *12*, 185−194.

Derivatization design example

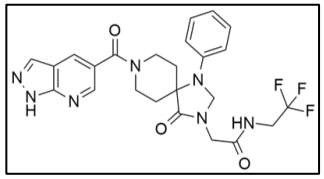
Setting	Value
Max_number_of_variations_with_same_linkage	1
Max_number_of_variations_with_varied_linkage	1
Max_number_of_total_modifications	1



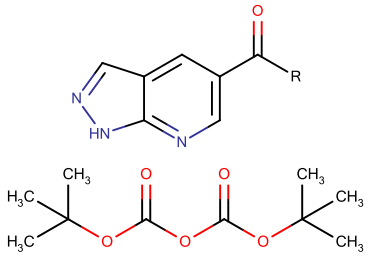
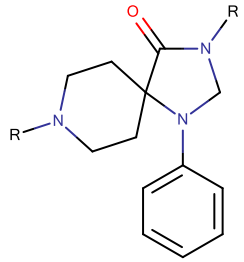
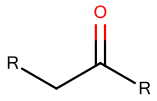
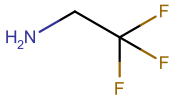
Similarity_step_0	0.9
Similarity_threshold	0.7
Similarity_threshold	0.7
Similarity_step_5	0.9

ACS Med. Chem. Lett. **2021**, *12*, 185−194.

Design results – with settings focused on the scaffold

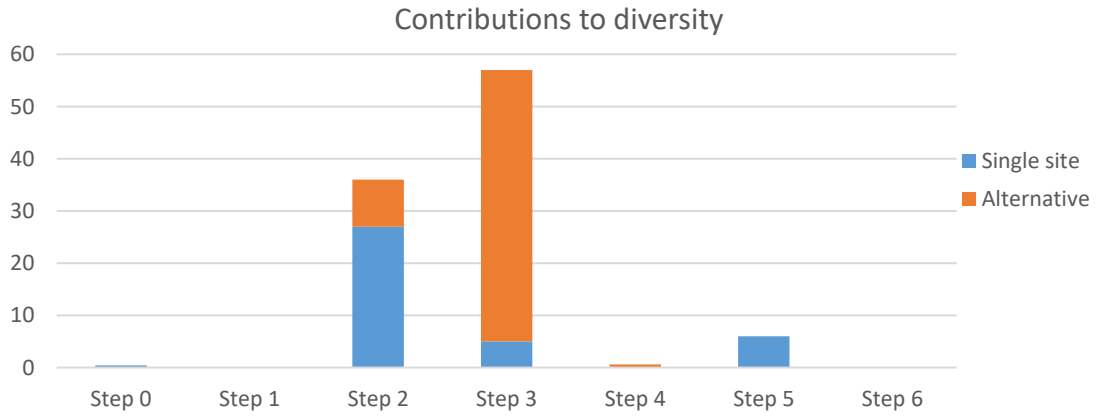


5,214 products

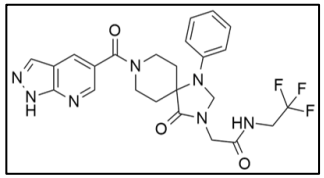
	Step 0	Step 1	Step 2	Step 3	Step 4	Step 5
						
	Similarity_step_0 0.9		Similarity_threshold 0.7	Similarity_threshold 0.7		Similarity_step_0 0.9
Site specific variations:	22 (0.4%)		1,876 (36%)	3,007 (57%)	30 (0.6%)	273 (6%)

<u>Positional (or reaction step) contribution to the total result set depends on:</u>

- Reagent (positional) similarity range set by user – primary driver
- Commercial (or custom uploaded) reagent diversity
- Reaction type – bimolecular adds much more than monomolecular

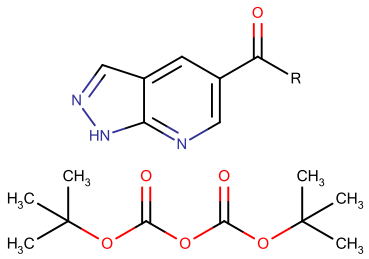


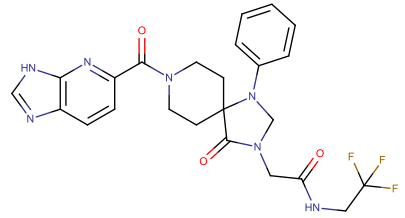
Example products

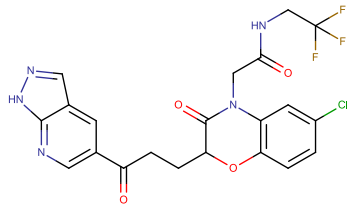


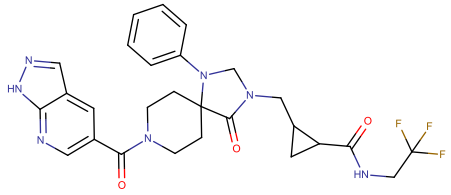
5,214 products

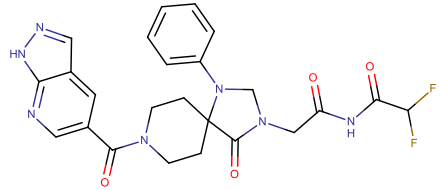
Step 0 | Step 1 | Step 2 | Step 3 | Step 4 | Step 5

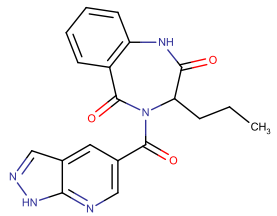


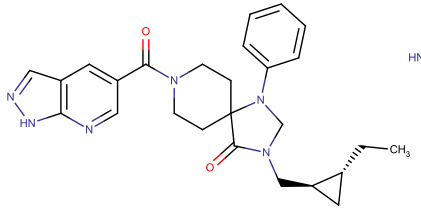


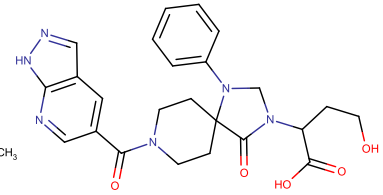




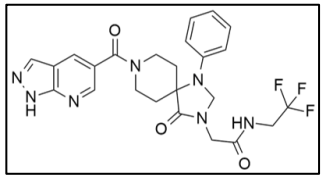






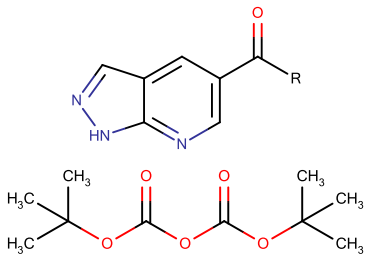
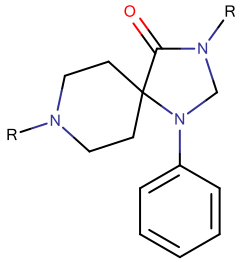
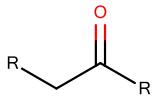
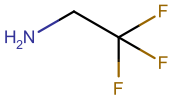


Design results – with equal similarity at each position

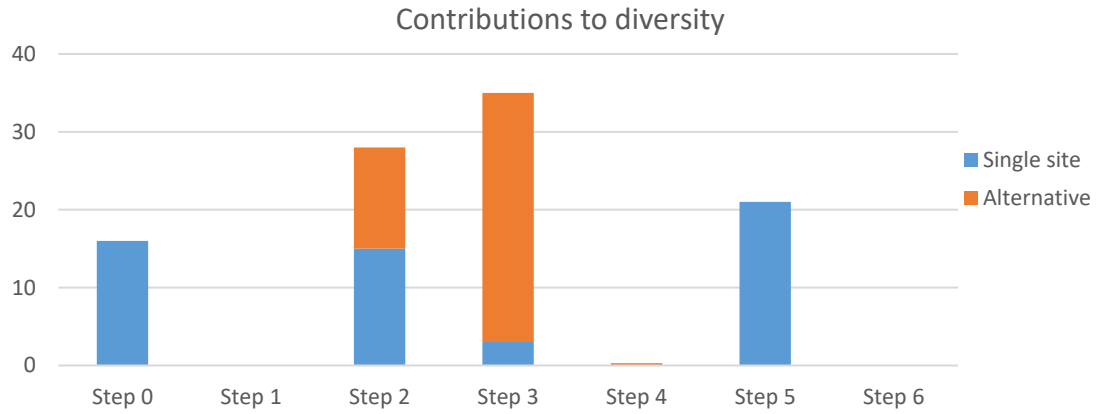


Similarity_threshold	0.7	1

9,264 products

	Step 0	Step 1	Step 2	Step 3	Step 4	Step 5
						
Site specific variations:	1,511 (16%)		2,564 (28%)	3,231 (35%)	30 (0.3%)	1928 (21%)

Positional (or reaction step) contribution to the total result set becomes well distributed



ACS Med. Chem. Lett. **2021**, *12*, 185−194.

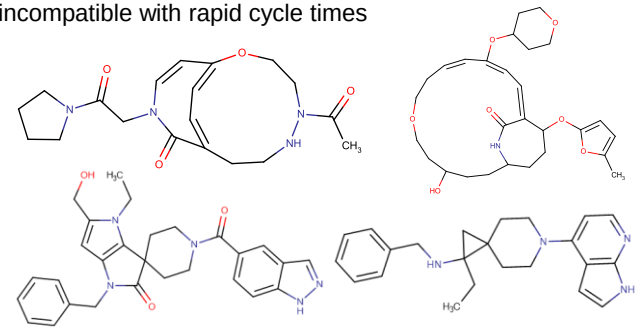
ChemPass

DDR1 case study: comparison of ChemPass Derivatization design (DD) to deep generative design by GAN (GD)

	GD	DD	Overlap	Indication (DD vs GD)
Size of training set	1,370	8	8	DD can be deployed in both early and later stages of a project, GD unlikely to have enough training data at early stages
Docked # cmpds	6,427	6,748	0	Comparable set size: <5% difference
Glide hit rate*	8%	34%		DD samples relevant chemical space much more so than GD. GD hit rate very low considering its large training set.
Best XP Glide score	-16.4	-17.7		DD optimizes leads, GD does not
Top 50 cmpds XP Glide score	-13.0	-15.5		DD set more populated with strong motifs
New scaffold chemotypes of interest	4	6	1	Both can produce novel motifs for new lead development directions
Synthetic feasibility**	mixed	high		Many challenging compounds in GD

* Hit rate: % compounds passed cutoff for XP docking after SP run

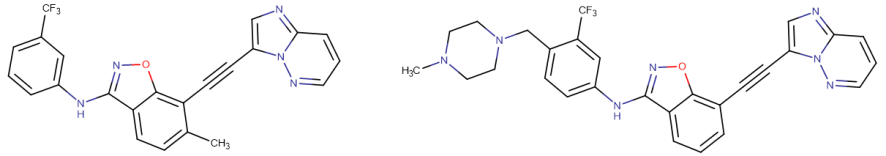
** GD (deep) generative design examples that are incompatible with rapid cycle times



10 synthetic steps 7 synthetic steps

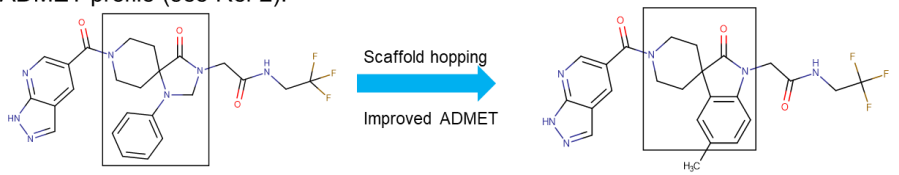
DERIVATIZATION DESIGN HIGHLIGHTS (simplest settings used for this study!)

Derivatization design discovered the key motif change present in the highlight molecule (cmpd 1) of the Nature article published by Zhavoronkov et al. (Ref 1):



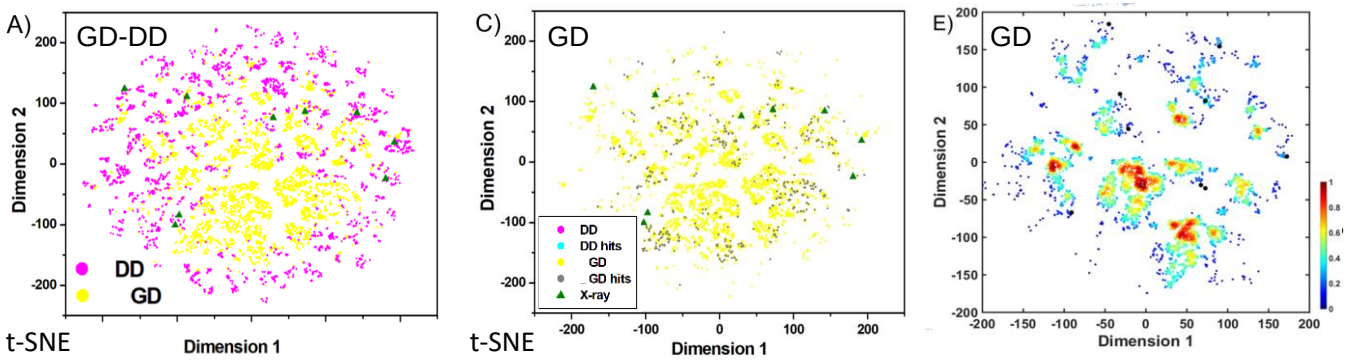
GD: Ref cmpd 1, Glide SP score: -13.9 (3ZOS), IC50: 10nM DD design: Glide SP score: -14.5 (3ZOS)

Derivatization design discovered the published improved spirocyclic motif possessing better ADMET profile (see Ref 2):



Ref cmpd 2.13, Glide SP score: -12.4 (6FEW) DD design: Glide SP score: -11.6 (6FEW)

DERIVATIZATION DESIGN samples more relevant chemical space for DDR1:



ChemPass Derivatization Design

- Is superior to the industry standard deep design in this study
  - Much higher hit rate
  - Better docking scores
  - More new motifs
- Can be used in early lead optimization
  - Needs no large training set
  - Can start from 1 hit/lead

Can effectively
- Optimize a lead, or
- Create new lead classes, or
- Do both at the same time

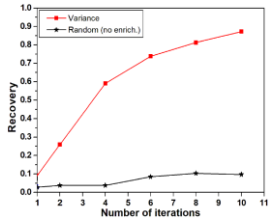
Produces synthetically feasible compounds
- 85% wet lab success rate for schemes with 3-6 steps

Additional in-house tools in AID Platform

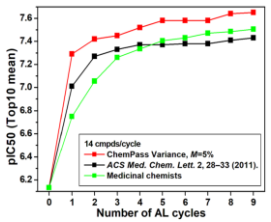
- Active learning (AL), desirability scoring, ML/DL models & automated analysis of docking results

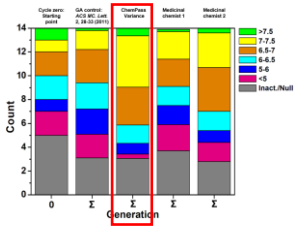
  - AL enhances throughput and speed

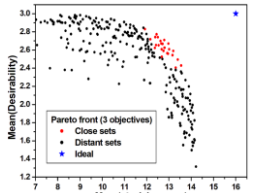
    Docking of 5-6% of the optimization idea set recovers 70-75% of "actives"



  - AL selection and desirability scoring outperforms GA & medicinal chemists







  - ML/DL models

PROTEINS: Structure, Function, and Bioinformatics 64:422–435 (2006)

Minimizing False Positives in Kinase Virtual Screens

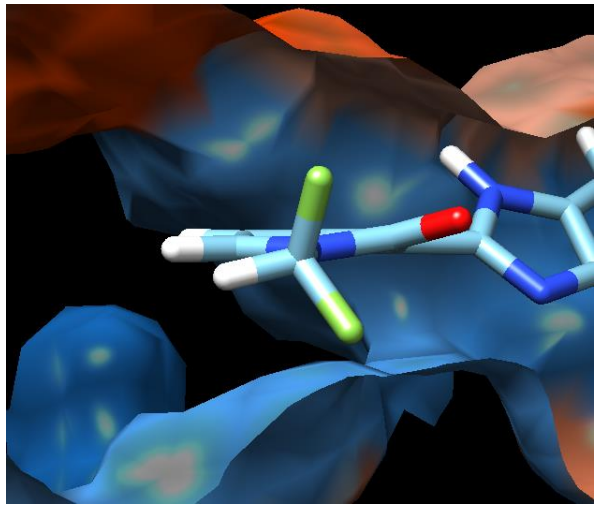
Emanuele Perola*
Vertex Pharmaceuticals, 130 Waverly Street, Cambridge, Massachusetts 02139

Ligand Strain Energy in Large Library Docking

https://doi.org/10.1101/2021.04.06.438722

Shuo Gu[1]†, Matthew S. Smith[1,2]†, Ying Yang[1], John J. Irwin[1], Brian K. Shoichet[1]*

- Conformational energy
- H-bonding (present and missing)
- Binding pose classification
- Multiple clash analyses
- Strain energies
- Ligand structural issues

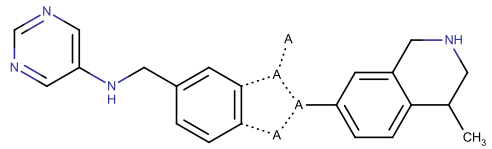


Significant distortion of aromatic planarity

Rapid elimination of decoys: 50-95% of poses get flagged or removed

Lead generation via scaffold hopping, Derivatization design and docking pose analyses

New DDR1 motif from Derivatization design case study:



Scaffold 1

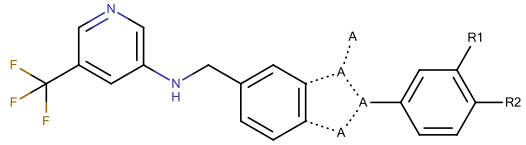
Glide Score: -13.1, DLE: 0.62, flagged poses!
(Ponatinib: GS: -15.4, DLE: 0.54, IC50: 9 nM)

Exploitation with

DERIVATIZATION DESIGN

Docking

20K analog space:
All docked molecules are flagged



Scaffold 1

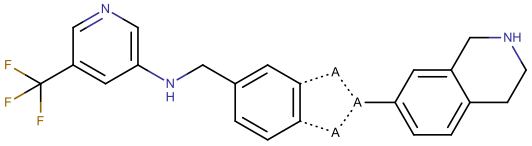
Best Glide Score: -16.7 with DLE: 0.65

Exploration with

1-CLICK SCAFFOLD DESIGN

Docking

39 relevant scaffold analogs with desired features:



Scaffold 2

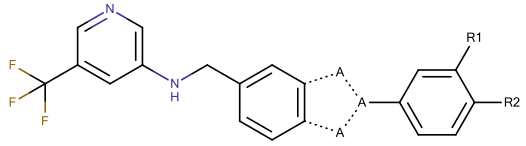
Glide SP Score: -12.5 with DLE: 0.55, no flags
Scaffold 1 control:
Glide SP Score: -14.0 with DLE: 0.60, flagged

Exploitation with

DERIVATIZATION DESIGN

Docking

30K analog space:
Many docked analogs have no flags

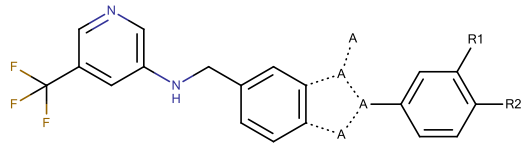


Scaffold 2

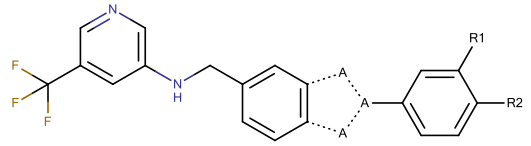
Best Glide Score: -16.9 with DLE: 0.68, no flags

Synthesis and assay

2 cmpds selected

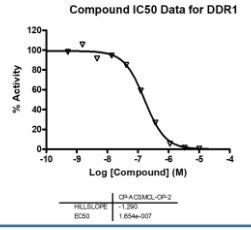


Scaffold 1, GS: -15.5 , flagged



Scaffold 2, GS: -14.9 , no flags

IC_{50}: >10uM



Exploration and exploitation in synthetically feasible lead analog space become simple, rapid and cost-effective processes

ChemPass

Summary

- *In silico* forward-synthesis technologies and user-friendly solutions have been developed for exploration and exploitation in synthesizable chemical space
- SynSpace users can solve different lead optimization tasks and get actionable results
- Derivatization design is a powerful generative design technique to cover relevant space around leads
- SynSpace forward-synthesis techniques can generate and optimize leads quickly and cost effectively
  - Compatible with wet and virtual cycles
  - Cycle time reduction
  - Cycle count reduction

Thank you for your attention!

Contact: Greg Makara

Phone: +36-30-389-9697

Email: gergely.makara@chempassltd.com